CMSC 657: Introduction to Quantum Information Processing
Lecture 22

Instructor: Daniel Gottesman

Fall 2024

1 Distance Measures

1.1 Fidelity

Recall that we defined fidelity between arbitrary density matrices ρ and σ to be

$$F(\rho, \sigma) = \operatorname{Tr} \sqrt{\rho^{1/2} \sigma \rho^{1/2}}. \tag{1}$$

Why is this the right measure?

Fidelity defined this way has a number of important properties:

Theorem 1. *Fidelity has the following properties:*

1. $0 \leq F(\rho, \sigma) \leq 1$
2. $\rho = \sigma \Leftrightarrow F(\rho, \sigma) = 1$
3. $F(\rho, \sigma) = 0$ *iff* ρ *and* σ *have support on orthogonal subspaces*
4. $F(\rho, \sigma) = F(\sigma, \rho)$
5. $F(U\rho U^\dagger, U\sigma U^\dagger) = F(\rho, \sigma)$ *for any unitary* U
6. $F(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \geq F(\rho, \sigma)$ *for any completely positive trace preserving map* $\mathcal{E}$
7. $F(\rho_A \otimes \rho_B, \sigma_A \otimes \sigma_B) = F(\rho_A, \sigma_A) F(\rho_B, \sigma_B)$
8. *Uhlmann's theorem:* $F(\rho, \sigma) = \max_{|\psi\rangle, |\phi\rangle} F(|\psi\rangle, |\phi\rangle)$*, where the maximum is taken over purifications* $|\psi\rangle$ *and* $|\phi\rangle$ *of* ρ *and* σ*, i.e., if* ρ *and* σ *are states on a Hilbert space* Q*, we purify with a system* R*, and* $\operatorname{Tr}_R |\psi\rangle\langle\psi| = \rho$*,* $\operatorname{Tr}_R |\phi\rangle\langle\phi| = \sigma$*.*

Uhlmann's theorem tells us that the definition of fidelity for mixed states is completely consistent with its definition for pure states, as the fidelity of two mixed states is the fidelity of the closest purifications of them. Dropping the reference system R is an example of a completely positive map, so for a general pair of purifications, the fidelity of the mixed states should be greater than or equal to that of the purifications; dropping information about R makes the states less distinguishable.

Two states are the same if their fidelity is 1 and completely distinguishable if the fidelity is 0. Sometimes people use the *infidelity*, $1 - F$, to reverse these numbers.

1.2 Trace Distance

Another useful notion of distance is the *trace distance*:

$$D(\rho, \sigma) = \frac{1}{2} \operatorname{Tr} |\rho - \sigma|. \tag{2}$$

This is the quantum version of the classical statistical distance between two probability distributions

$$D(\{p_i\}, \{q_i\}) = \frac{1}{2} \sum_i |p_i - q_i|. \tag{3}$$

When ρ and σ are simultaneously diagonalizable, we get exactly the classical statistical distance between the probability distributions given by their eigenvalues. In general, if we diagonalize $\rho - \sigma$, it has eigenvalues λ_i, and

$$D(\rho, \sigma) = \sum_i |\lambda_i|. \tag{4}$$

Theorem 2. *The trace distance has the following properties:*

1. $0 \leq D(\rho, \sigma) \leq 1$
2. $D(\rho, \sigma) = 0 \Leftrightarrow \rho = \sigma$
3. $D(\rho, \sigma) = D(\sigma, \rho)$
4. *Triangle inequality:* $D(\rho, \sigma) \leq D(\rho, \eta) + D(\eta, \sigma)$
5. $D(U\rho U^\dagger, U\sigma U^\dagger) = D(\rho, \sigma)$ *for any unitary* U
6. $D(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \leq D(\rho, \sigma)$ *for any completely positive trace preserving map* $\mathcal{E}$.
7. $D(\rho \otimes \eta, \sigma \otimes \eta) = D(\rho, \sigma)$.

For trace distance, 0 means the states are the same and 1 means that they are completely distinguishable. Thus, a CPTP map which loses information should make the states less distinguishable and D gets smaller.

Note that trace distance is *not* multiplicative: $D(\rho_A \otimes \rho_B, \sigma_A \otimes \sigma_B) \neq D(\rho_A, \sigma_A) D(\rho_B, \sigma_B)$ in general.

The operational meaning of classical statistical distance is that it gives the probability, given one sample from one of the two distributions, of correctly guessing which distribution it is. Trace distance has a very similar operational meaning:

Theorem 3. $D(\rho, \sigma)$ *is the maximum over POVMs* $\{M_a\}$ *of the classical statistical distance between the two probability distributions* $\operatorname{Tr}(M_a \rho)$ *and* $\operatorname{Tr}(M_a \sigma)$. *Thus,* $D(\rho, \sigma)$ *gives the probability of correctly guessing which density matrix one has after making the optimal measurement:*

$$\textit{Prob.(correct guess)} = \frac{1}{2}(1 + D(\rho, \sigma)). \tag{5}$$

(Assuming a priori 50% *of either.)*

Proof. Let us diagonalize $\rho - \sigma$. It has trace 0, so it has a subspace S_+ of positive eigenvalues and another orthogonal subspace S_- of negative eigenvalues (with zero eigenvalues assigned to either). Let $\Pi_\pm$ be the projector on $S_\pm$ and let $P_\pm = \Pi_\pm(\rho - \sigma)\Pi_\pm$. I claim that the optimal POVM is a projective measurement $\{\Pi_+, \Pi_-\}$. Certainly, for this measurement,

$$\begin{aligned} D(\{p_i\}, \{q_i\}) =& \frac{1}{2} \left[|\operatorname{Tr}(\Pi_+ \rho) - \operatorname{Tr}(\Pi_+ \sigma)| + |\operatorname{Tr}(\Pi_- \rho) - \operatorname{Tr}(\Pi_- \sigma)| \right] && (6) \\ =& \frac{1}{2} \left[|\operatorname{Tr} \Pi_+ (\rho - \sigma)| + |\operatorname{Tr} \Pi_- (\rho - \sigma)| \right] && (7) \\ =& \frac{1}{2} \left[\operatorname{Tr} \Pi_+ (\rho - \sigma) - \operatorname{Tr} \Pi_- (\rho - \sigma) \right] && (8) \\ =& \frac{1}{2} \operatorname{Tr} |\rho - \sigma|, && (9) \end{aligned}$$

with the last two lines following because $\Pi_\pm$ is a projector onto $S_\pm$.

This POVM is actually the optimal one. If we have any other POVM $\{M_a\}$, then we again have

$$D(\{p_i\},\{q_i\}) = \frac{1}{2}\sum_a |\operatorname{Tr}(M_a(\rho-\sigma))|. \tag{10}$$

If we write $\rho - \sigma = P_+ + P_-$ (which follows in a basis where $\rho-\sigma$ is diagonal), then

$$\operatorname{Tr}(M_a(\rho-\sigma)) = \operatorname{Tr} M_a P_+ + \operatorname{Tr} M_a P_-. \tag{11}$$

Now, P_+ is a positive semidefinite operator, as is $-P_-$ (since the eigenvalues of P_- are all negative or 0). Thus,

$$|\operatorname{Tr}(M_a(\rho-\sigma))| \le |\operatorname{Tr}(M_a P_+)| + |\operatorname{Tr}(M_a P_-)| = \operatorname{Tr}(M_a P_+) - \operatorname{Tr}(M_a P_-). \tag{12}$$

But

$$\operatorname{Tr}\Pi_\pm(\rho-\sigma) = \operatorname{Tr} P_\pm, \tag{13}$$

so

$$\begin{aligned} D(\{p_i\},\{q_i\}) \le& \frac{1}{2}\sum_a (\operatorname{Tr} M_a P_+ - \operatorname{Tr} M_a P_-) && (14)\\ =& \frac{1}{2}\left[\operatorname{Tr}(\sum_a M_a)P_+ - \operatorname{Tr}(\sum_a M_a)P_-\right] && (15)\\ =& \frac{1}{2}(\operatorname{Tr} P_+ - \operatorname{Tr} P_-) && (16)\\ =& D(\rho,\sigma), && (17)\end{aligned}$$

as we calculated above. □

Theorem 4. *The fidelity and trace distance between two states satisfy the following relationship:*

$$1 - F(\rho,\sigma) \le D(\rho,\sigma) \le \sqrt{1-F(\rho,\sigma)^2} \tag{18}$$

1.3 Operator Norms

We are also interested in ways of saying two unitaries or two CPTP maps are close to each other. There are a number of different distance measures between linear maps as well, and again there is not a single "right" one to use in all circumstances. In this case, it is little more common to consider norms, which measure the size of an operator, rather than distances between two operators. However, given a norm $\|\cdot\|$, we can get a distance between A and B by subtracting the two operators $\|A-B\|$.

One common operator norm (which we saw earlier) goes by a few names, such as the sup norm or the infinity norm. The sup norm is applied to linear operators acting on the Hilbert space, and it is defined as

$$\|A\|_{\sup} = \sup_{|\psi\rangle} \|A|\psi\rangle\|. \tag{19}$$

Here, the supremum is taken over normalized states. The sup norm is the maximum amount the operator can stretch a vector and is equal to the largest eigenvalue of the matrix.

We also need norms for linear maps on density matrices. These can give norms for linear operators on the Hilbert space as well, of course, since they also transform density matrices, but apply to other things as well. Note that we don't want to restrict to trace preserving maps or even to completely positive maps since we want to apply the norm to differences to give a distance, and differences of CPTP maps are not CPTP.

A first attempt is to do something analogous to the sup norm but using density matrices. This actually gives us a couple of options because we have multiple sensible measures for the norm of a density matrix. One common choice is to use the trace distance, or 1-norm, for the density matrix:

$$\|\mathcal{C}\|_1 = \sup_\rho \frac{\|\mathcal{C}(\rho)\|_1}{\|\rho\|_1}. \tag{20}$$

Here, the supremum is taken over positive matrices ρ. In this formula, we divide by the trace instead of restricting to $\operatorname{Tr}\rho = 1$, but you could do it either way.

One problem with this formula is that

$$\|\mathcal{C} \otimes I\|_1 \neq \|\mathcal{C}\|_1 \tag{21}$$

in general; $\|\mathcal{C} \otimes I\|_1$ can be bigger. This occurs because there can be an entangled state on the tensor product system which gives a bigger value for the supremum than any tensor product state. This means that the 1-norm of operator maps doesn't fully capture the power of the map on the size of entangled states. Moreover, it leads to a problem if we want to talk about the size of a particular map, say an error, since the size of the map can depend on what other qubits are around even though they are not directly involved in the map.

The solution is to include the extra degrees of freedom in the supremum:

$$\|\mathcal{C}\|_\diamond = \sup_\rho \frac{\|(\mathcal{C} \otimes I)(\rho)\|_1}{\|\rho\|_1}. \tag{22}$$

This *diamond norm* now looks at the supremum over operators ρ on a tensor product space including the domain of the original map $\mathcal{C}$ and a reference system. The value you get can depend on the size of the reference system, but luckily it saturates when the reference system is the same size as the original space (the domain of $\mathcal{C}$). This is basically because a reference system of the same size as the system allows the maximal amount of entanglement with the system.

The diamond norm then has the nice property that $\|\mathcal{C} \otimes \mathcal{D}\|_\diamond = \|\mathcal{C}\|_\diamond \|\mathcal{D}\|_\diamond$.

It is also possible to make an analogous set of operator norms based on the fidelity rather than the trace distance. Another thing that sometimes varies is to use averages over states instead of worst-case states. All of these variations have their place and applications.